Girardi *et al. Safety in Health* 2015, **1**:6
http://www.safetyinhealth.com/content/1/1/6

SAFETY IN HEALTH

RESEARCH ARTICLE Open Access

An ontology-based clinical data warehouse for scientific research

Dominic Girardi*, Johannes Dirnberger and Michael Giretzlehner

Abstract

Medical research but also quality management is based upon medical data. The integration, validation, processing, and exploration of this data is known to be a technical obstacle for researching medical domain experts and a major pitfall to (bio-)medical research projects. To overcome this pitfall and actively support the medical domain expert in these tasks, we present an ontology-based clinical data warehouse for scientific research. It is completely generic and adapts itself at run-time to the current domain-ontology, which can be freely defined by the domain expert and describes the actual field of research. The whole system adapts is appearance and behavior to this central ontology and appears to the user like a custom made solution. Furthermore, the elaborate structural meta-information from the ontology is used to actively support the user in tasks that usually require profound IT knowledge, such as defining complex search queries or data quality constraints, or applying advanced data visualization algorithms to the data. The proposed warehouse supports the domain expert trough the whole process of knowledge discovery from data integration to exploration.

Keywords: Clinical data warehouse, Research data integration, Exploratory data analysis

Background

The process of knowledge discovery in databases (KDD) is a well known and commonly agreed process in the field of computer science. Cios et al. [1] define this process '*as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*'. The process, which usually contains steps as understanding the problem and data, data preparation, and data mining, is mostly performed by so called data scientist. The (medical) domain expert is seen in a super-vising, consulting and customer role. The KDD in the field of medicine differs significantly from this description since the role of the domain expert switches from the external supervisor to the main actor of the process, which is mostly caused by the complexity of the research domain [2]. These domain experts are now confronted with a large amount of highly complex, heterogeneous, semi-structured research data of often poor quality. The handling, processing and analysis of this data is known to be a major technical obstacle for (bio-)medical research projects [3], including tasks as:

- Data integration from numerous, heterogeneous resources
- Adaptions and extensions to the domain data structures and subsequent adaptions of the whole infrastructure
- Data selection by complex criteria out of databases
- Assuring and checking data quality
- Derivation of new attributes out of already existing ones (feature generation)
- Data pre-processing, normalization and transformation for the subsequent application of analysis or visualization algorithms

The only appropriate way to gain a structured and valid data pool for scientific medical research — which is handleable by medical domain experts — is the implementation of a clinical data warehouse, like Prokosch and Ganslandt [4] suggest. 'A data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of managements Decision support process' [5]. In a clinical data warehouse all relevant data is accumulated and stored in a highly structured way to allow the analysis of the data. A clinical data warehouse which is able to actively support the user in all these tasks

*Correspondence: dominic.girardi@risc.uni-linz.ac.at
Research Unit Medical Informatics, RISC Software GmbH - Johannes Kepler University Linz, Softwarepark 35, Hagenberg, Austria

needs to be strongly customized to the users needs and the underlying data structures. These data structure of a clinical research data warehouse are designed for and adapted to the corresponding research domain. Consequently, a warehouse designed and implemented for a certain research domain is hardly reusable for a different domain. This requires the individual development of a clinical data warehouse which results in high costs and development effort and duration.

We present a domain-independent, ontology-centered clinical data warehouse for medical research. It is based upon a generic meta data-model and is able to store the current domain ontology (formal description of the actual research domain) as well as the corresponding research data. The whole warehouse is implemented at a higher level of abstraction and derives its manifestation and behavior from the actual domain ontology at run-time. Just by modeling the domain ontology, the whole warehouse, including electronic data interfaces, web portal, search forms, data tables, etc. is customized for the actual usage. Furthermore, the stored domain ontology is used, to actively support the user in tasks that usually require profound IT knowledge or support, such as querying data sets of interest by non-trivial search conditions, data aggregation, feature generation for subsequent data analysis, data pre-processing, and the application of advanced data visualization methods. In this way, a system could be developed which is able to process data of arbitrary structure and at the same time behaves for the user like it was individually developed for the actual application.

Related research

The idea of using meta models to automatically create parts (data access layer, user interfaces) of data intensive software systems is widely established method in model driven engineering (MDE) [6]. However, the MDE approach in general or concrete realizations like the meta model-based approach for automatic user interface generation by da Cruz et al. [7] — to give an example — are used by software engineers to create source code or source code skeletons at development time. Cruz et al. use extended UML meta models to describe the domain data model and use cases and derive source code from it - source code that needs to be compiled. That's the main difference between our approach and MDE: Our system derives the structure of the user interface from the meta model at run-time. There is no source code generation. Changes to the domain data model have immediate effect to the user interface, without any compiling or even restart of the application. From this perspective our system is related to the Margitte system by Renggli et al. [8]. While the Margitte system is a general purpose framework based on a self-describing meta-model, our system is based upon a meta-entity-relationship model (see Section Data model) - stored in relational database - and clearly focused and specialized on scientific data-acquisition and -processing considering the biomedical research as a main user and administrator. Asides from the automatically created web interface it offers a corresponding software tool to handle, pre-process (using the expression engine, described in this paper) and subsequently analyze the collected data. So, the main idea of our approach is the generation of a highly flexible data acquisition and management system due to the interpretation of meta data models at run-time. There is a close relation to ontology based systems. Zavaliy and Nikolski [9] describe the use of an ontology for data acquisition, motivated by the demand of adaptive data structures. They used an OWL ontology to model their domain, which consists of four concepts (*Person*, *Hospital*, *Diagnosis* and *Medication*). There is no information given on user interface generation. Asides from this work, it was very hard to find any publications on absolutely generic ontology based data acquisition systems. In most publications ontologies are used for information extraction from text [10] or to enrich the existing data with structural and semantic information or to build a cross-institutional knowledge base. In [11] the authors describe the usage of ontologies for inter-hospital data extraction to improve patient care and safety by analyzing hospital activities, procedures, and policies. Here, the ontology is strongly connected to the project. In [12] e.g. the authors describe an ontology based system for extracting research relevant data out of heterogeneous hospital information systems. Here again, the purpose is to find a superior common data model to compare data of different medical institutions. The most comparable work was published in 2014 by [13] Lozano et al., who also present an ontology-based system, called OWLing. Their intention is comparable to the abovementioned, but their implementation is completely based upon the web ontology-language OWL. Our approach differs from OWLing and many other in the aspect that it covers the whole knowledge discovery process from data modeling to exploratory data analysis in one integrative platform, while other projects focus on parts of it.

From a functional point of view, our system is closely related to other electronic data capture systems for medical purposes such as Catalyst Web Tools, OpenClinica and REDCap [14]. These systems are very complex and offer additional features for study management and planning.

Methods

Theory

The idea of the system is based upon a modification of the process model for knowledge discovery. The process-step *data modeling* can hardly be found in common definitions of the KDD — see comparison in [15], Table one on page 6.

This is due to the fact that these process definitions are mostly made for data, which is already available and steady concerning its structure. By introducing this step into to the process model and enabling the researcher (domain expert) to model the domain of interest the existing bias of "exploring what's available" can be overcome.

The domain experts defines what data is necessary for the current research project, without bothering in a first step where this data comes from. Based on this domain data model, which is called the domain ontology, the data warehouse derives its actual structures and behavior, including electronic data import interfaces and web interfaces for manual data input. These two data interfaces can now be used to integrate data from numerous, heterogeneous sources into the data warehouse. Especially electronically available data in the field of medical or clinical research is often spread over several systems (hospital information systems, laboratory information systems, surgery documentation systems, etc.), often semi-structured and heterogeneous concerning data types — ranging from numeric to categorical data.

We use a generic meta data-model which is able to store the actual domain-ontology and the corresponding data. The Object Management Group OMG [16] defines four levels (M_0 - M_3) of meta modeling. Each model at level M_i is an instance of a model at level M_{i+1}. Level M_0 contains the real world transactional data. Each object at M_0 is an instance of a model defined in M_1, which is called the model layer. Each model at level M_1 can be seen as an instance of a meta model at level M_2 - the meta model layer. Level M_3 contains meta meta models.

Conventional data storage systems use M_1 data models, which directly describe the field of application. M_2 meta models describe M_1 data models. We developed a meta model which is able to store M_1 data models as well as their M_0 transactional data. This allows replacing the conventional M_1 data model by the M_2 meta model, which — of course — requires the whole application that is built upon this M_2 model to be able to process this kind of abstract data model.

Data model

Since most M_1 data models are relational models stored in a relations database, they can be described using an Entity Relationship model (ER model), which was introduced by Peter Chen in 1976 [17]. Consequently a meta model, which is able to describe an ER model can be used to store data structures that are stored in a relational database. For a more detailed explanation of the used data model and the project's relation to OWL ontologies, the reader is referred to the article *An Ontology-Based Data Acquisition Infrastructure* [18].

Usage

To prepare the generic system for the actual application, the domain-specific ontology must be modeled, in order to store the structural information into the meta data-model. Therefore, the user defines what data entities exist in his domain (e.g. Patient, Disease, Treatment, etc.), what kind of attributes they have (e.g. Patient.DateOfBirth, Disease.DiagnoseDate), and in what kind of relationship they are (e.g. a Patient can have numerous diseases). Furthermore, data validity rules, ranging from simple numeric ranges to complex expressions representing medical domain knowledge, are also part of the domain ontology. These definitions can be done using corresponding wizards and web forms and don't require any programming. Based on this model definition, the whole data warehouse (data management environment, web input forms, search forms, overview tables, data import and export interfaces, etc.) is created automatically at run-time. The ontology can be changed at any time of the research project. So, researchers are independent from their software vendors and can adapt their system to their needs at any time, whereas the system prevents the user from changes that would cause data loss. Due to the user interface generation at run-time, all changes are propagated throughout the whole system immediately. Furthermore, by using the stored meta information about the domain ontology, the system is able support the user in data management, data processing, definition of complex validity rules and and data analysis and visualization.

Results

Ontology-based components

The main challenge is to gap the discrepancy between an generic, general purpose system and a system which is able to actively support the user in dealing with the current domain ontology. The warehouse appears to the user like a system which was individually developed for the current application, while it is absolutely generic and able to process almost any data structure. Consequently, all relevant components of the infrastructure are implemented at a higher level of abstraction with a strong linkage to the current domain ontology.

Ontology-based data integration

The integration of electronically available and structured data is the main purpose of a clinical data warehouse. This requires a generic, ontology-based data import interface which is able to map from arbitrary input sources onto the current domain ontology. Reading data from external sources, transforming it and loading it into target data bases is a classical use case for ETL (Extraction-Transform-Load) processes. Since there are numerous solutions available for performing ETL, it is possible to leave the extraction and transformation part up to those

solutions. The crucial part is the loading which bridges the gap between the flat-table based ETL data structures and the arbitrarily complex domain ontology. So there is a flat data table, which comes from the ETL process data stream on the one hand and the domain ontology on the other. It is now up to the user to define which columns from the input table map onto which attributes of the actual domain ontology. The ontology-based data integration module validates the user defined mappings and ensures that they are consistent within themselves and in accordance with the current domain ontology. Once the mapping is defined and valid, the mapping algorithm maps each row if the data input-stream onto the ontology structures and updates and/or creates the corresponding data records.

Ontology-based user interfaces

User interfaces for data acquisition and presentation are a central aspect of the system. Since the system is generic and the data structures are unknown, all user interfaces are created at run-time based on the actual state of the domain ontology. There are three basic display modes of data records:

1. **The display of a number of records from the same entity type:** When records that are instance of the same entity (this one will be called the base entity of the table) are displayed at once, the well established form of a table is chosen. This is a very intuitive and expected way of displaying numerous data records that show the same structure. The structure of the table is derived from the domain ontology meta-information of the base entity.
2. **The display of a number of records, from different entity types that belong to one data set in form of a record tree:** The record tree display allows to view the current record in the context of the whole data set. The tree representation is an intuitive and well-known way to present hierarchically organized data. Due to the restriction to the data model to only allow tree-like structures this type of presentation is always possible. The actual tree representation of a current record is an instance of the hierarchical structure represented by the domain ontology.
3. **The display of one single record:** Single records are presented in a form-like display. Comparable to the multi-record table where the entity's attributes are used to create the table columns, here a row is created for each attribute. Single record displays can be shown in read-only mode or in edit mode. When opened in edit-mode, the whole data set of the opened record is locked for all other user in order to prevent concurrent data manipulations. Each row consists of two elements: Firstly, a label, which is the attribute's name. Secondly, and a value representation, which is — dependently of the operation mode — either a read-only representation of the actual value or a data type dependent input field.

The concept of the ontology-based creation of a medical data acquisition system was published in [19].

Ontology-guided expression engine

Whenever data is stored in a structured way - eg. databases, ontologies, XML - an expression language (SQL for databases, SPARQL for OWL ontologies, xQuery for XML) is provided to define complex search queries and data validity rules. Moreover, for data analysis it is often necessary to use expressions to derive the features to analyze out of the already existing data. Especially in the field of medicine, rather changes and differences (functions of data values) carry information than the data itself [20]. Usually the usage of these expression languages is reserved for IT experts who are familiar with the grammar of the language and the internal data structures of the application.

An expression in the field of computer science is defined as any piece of program code in a high-level language which, when (if) its execution terminates, returns a value. In most programming languages, expressions consist of constants, variables, operators, functions [21]. This definition already gives a hint, why the usage of expressions in data handling is usually reserved for IT experts. At first, an expression is a piece of programming code. Secondly, the code must be valid in terms of the grammar of a high-level language in order to execute. Thirdly the definitions implies that the expression language includes an amount of operators and functions. In order to use the expression language the user should be familiar with at least some of those operators. Additionally, when the expressions are executed on a data structure, they must be correct in terms of this structure as well; concerning compatibility of variables (from the data structure) and operators (part of the grammar).

In order to support the medical researcher in modeling his domain knowledge as grammatically correct expressions, an ontology-guided expression editor was developed. This editor interweaves grammatical and structural meta-information to guide the user through the expression definition.The expression editor is not a text-based approach, but a tree-based solution. Each node in the tree is an operator of the expression language and the user can interact with these nodes using the right-click context menu. If the selected node has any possible successors, according to the grammar, then these successors are offered in this context menu. This way, only grammatically

correct successors are offered to the user, who is not able to create a grammatically wrong expression - except the possibility to create an incomplete expression. So, there is no need to memorize big amounts of operators and their input parameters. Furthermore, the tree offers comfortable operations like drag and drop and expand and collapse. There are, of course, expression operators that access the data and yield, e.g., all numeric fields of a certain class. If there are no such fields in the current class, the operator is not offered in the context menu, although it would be correct in terms of grammar. Furthermore, there are a number of expression operators which allow the user to traverse along the relations that are stored in the ontology and connect the classes with each other. So, data from different classes can be combined when defining expressions. The expressions can be used in three ways: for the definition of complex search criteria, the generation of new features for data analysis, and for the definition of data validity rules. For a more detailed description of the ontology-guided expression engine, the reader is referred to [22].

Ontology-supported data exploration

In contrast to statistical approaches aimed at testing specific hypotheses, Exploratory Data Analysis (EDA) is a quantitative tradition that seeks to help researchers understand data when little or no statistical hypotheses exist [23]. Its aim is to provide an overview of the data and an idea of correlations and patterns that might be hidden within the data. Exploratory analysis often comes along with data cleaning, when the exploration yields outliers and irregularities. EDA is often supported by visual means, being known under the buzzword visual analytics, which is defined as the science of analytical reasoning facilitated by interactive visual interfaces [24]. Before EDA can be performed on data, which is stored in a non-trivial data structure and in a database, a number of data preprocessing steps need to be performed including defining database queries to retrieve data of interest, data aggregation, combination, transformation and normalization. Many of these steps, especially those which require programming or database knowledge, are hardly manageable for medical researchers. This also applies to the application of sophisticated analysis and visualization methods like data mining and clustering algorithms, and complex visualization methods.

The domain ontology provides the meta-information to support the medical researcher in all these aspects. The ontology-guided expression engine is used to query and prepare the data sets of interest. The data-preparation for the actual analysis method of choice can then be automatized, using the ontological meta-information. In the course of a feasibility study three non-trivial visualization were implemented: A Self-Organizing or Kohonen Map [25], a Parallel Coordinate Visualization [26] (see Figure 1), and a non-linear dimension reduction Sammon's Mapping [27] — see [28,29]. All these visualization show high dimensional data on a two-dimensional display and allow the visual identification of clusters and patterns. They can be applied to any data set of interest, independently from the current domain ontology.

System modules

From the user's perspective the system presents itself in three modules, which integrate the above-mentioned components.

1. **Management tool:** The Management Tool is the main point of interaction for the research project leader. It allows the modeling and maintenance of the current domain ontology (see Figure 2), as well as data processing, data validation (see Figure 3, and exploratory data analysis.
2. **Data interface:** The data interface is a plug-in to the well established open source ETL (Extraction-Transform-Load) suite Kettle, by Pentaho. Kettle allows the integration of numerous data sources and enables the user to graphically model his ETL process. For the final step, the data integration into the data warehouse the concept of the above-described Ontology-based Data Integration was implemented.
3. **Web interface:** The web interface is an automatically created web portal, which allows the users — depending on their privileges – to enter, view and edit existing data records. It is usually used to manually complement the electronically imported data with information that was not available electronically (e.g. hand-written care instructions, fever graphs, ect.).

Applications

The presented clinical data warehouse is already applied in a number of medical research and benchmarking projects with a high variety concerning their domains and domain ontologies. One of the earliest applications is a web-registry for cerebral aneurysms run by the department for Radiology of the Landesnervenklinik Linz Wagner Jauregg. In the same clinic, an internal database for outcome studies on neuro-surgical interventions was established. The children's hospital of the country Upper Austria in Linz is using the warehouse for a biometric study for children and young adults. The department for Process Management in Health Care of the university of applied sciences in Steyr applies this infrastructure for systematic benchmarking of surgical treatments over eight different clinics. Furthermore, the proposed system is already used by a Austrian major league soccer and ice-hockey club for sports medical documentation and analysis.

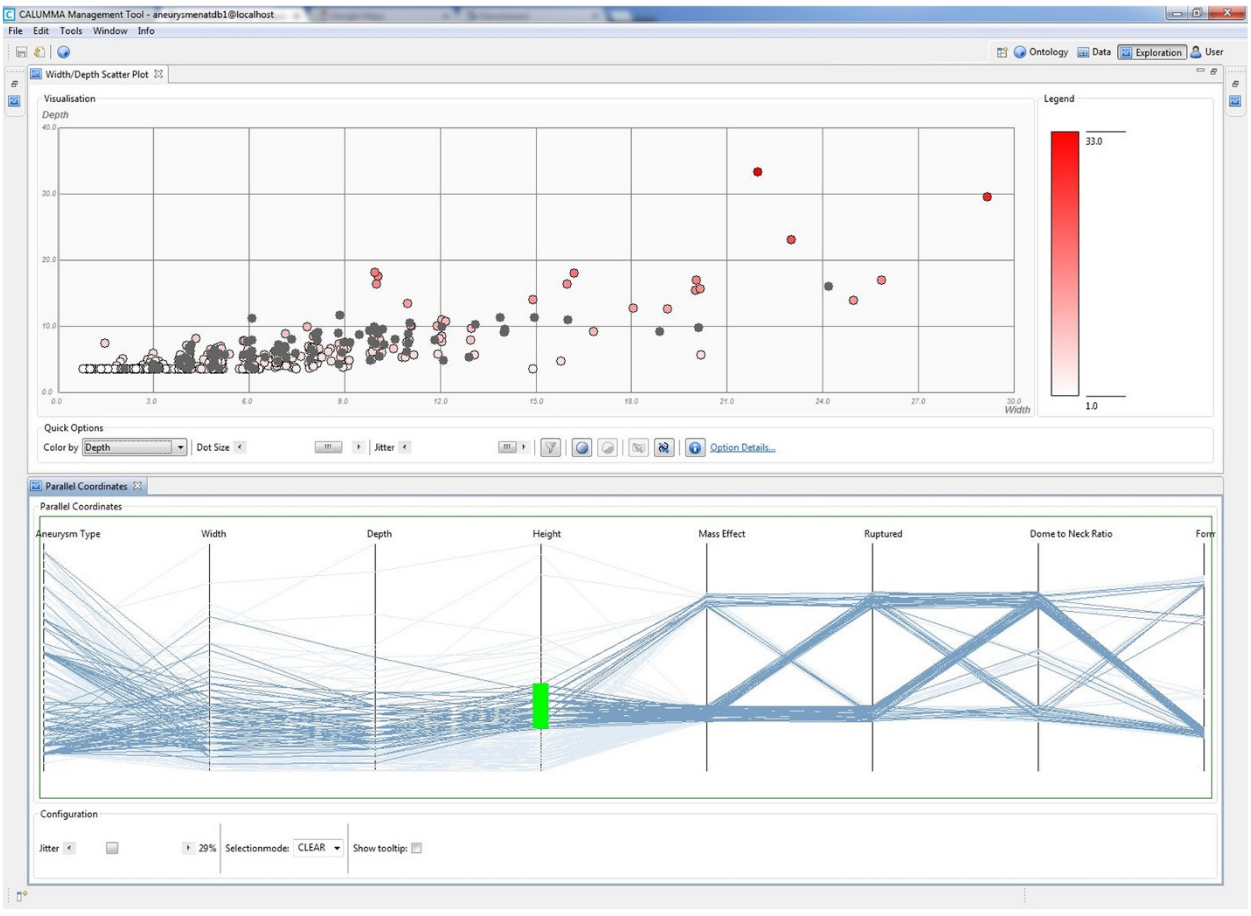

Figure 1 The visual analytics perspective showing a medical data set. While the upper section shows a conventional scatter plot of two numeric dimensions, the lower area contains a parallel coordinate system showing eight dimensions of the data set. Selections in one of the visualizations are highlighted in the other one to allow the user to recognize patterns and correlations across the visualization methods.

Discussion

The presented system is based on a critical questioning of the knowledge discovery process in the field of medical research. The proposed approach of an ontology-centered data warehouse has already proven suitable and practicable to solve the problems that arise to medical domain experts when they are confronted with their research data. The set of features and their implementation showed to be sufficient to cover most relevant aspects of a data-intense research projects. Although the system allows the integration of specialized source code to customize the warehouse beyond its configuration possibilities, this feature was hardly used so far.

While we could show that the system is able to support the medical domain expert in their work with an ontology-centered system, the initial configuration of this ontology turned out to be difficult task for non-IT users. When confronted with a completely empty initial configuration medical researchers did not know how to build a proper domain ontology from scratch. They were not aware of the consequences their modeling has to all subsequent tasks (acquisition, integration, exploration, etc.). So in this aspect, the system falls behind the aspiration to be handleable completely without expert IT knowledge. So far, this drawback is compensated by support and consulting in the data modeling phase. However, it's also an option to provide standard ontologies — reference models — which can be extended and adapted by the users. Experiences showed that the obstacle to extend and adapt an existing ontology is lower that building and ontology from scratch.

Although the system is able to capture and store unstructured data (free text) in text fields and text boxes, it is not able to process this data any further in terms of exploration. Only structured (categorical) or numeric data (including date data types and Boolean data type) can be processed. The user is encouraged to avoid unstructured data and create his own categorical enumerations instead, which are part of the the domain ontology and can be defined freely. So free-text can be avoided by offering corresponding categorical enumerations to be selected.

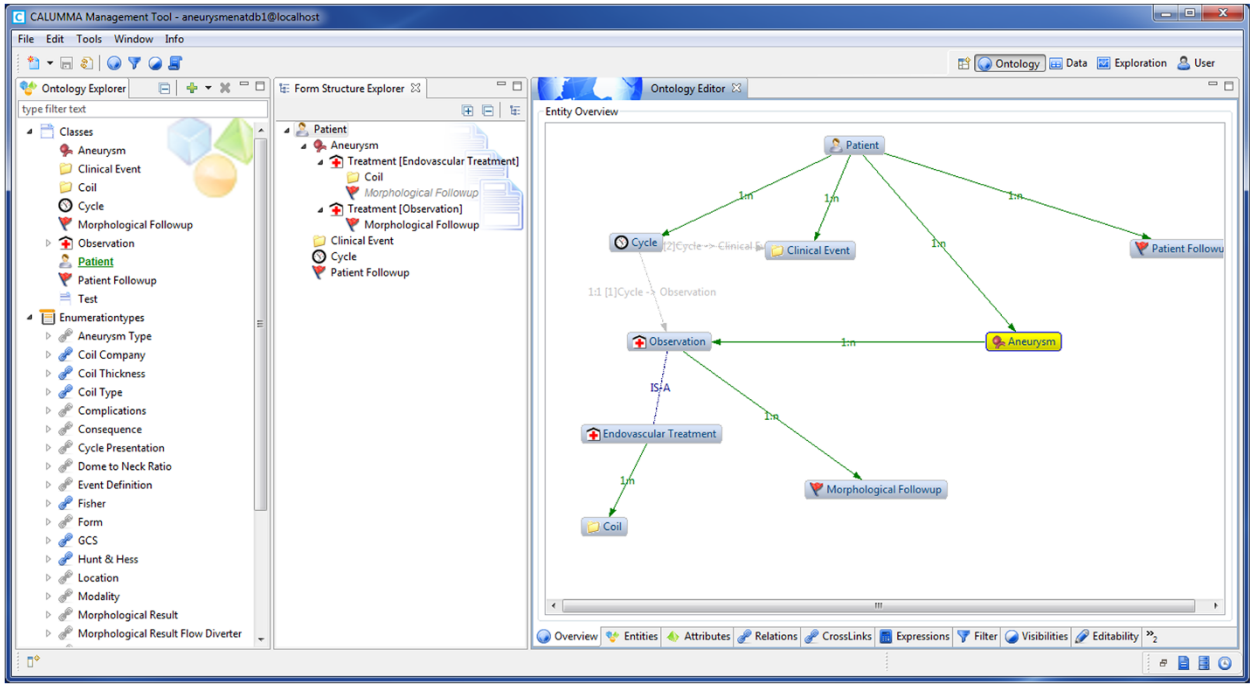

Figure 2 An example domain ontology shown in the ontology modeler of the Management Tool.

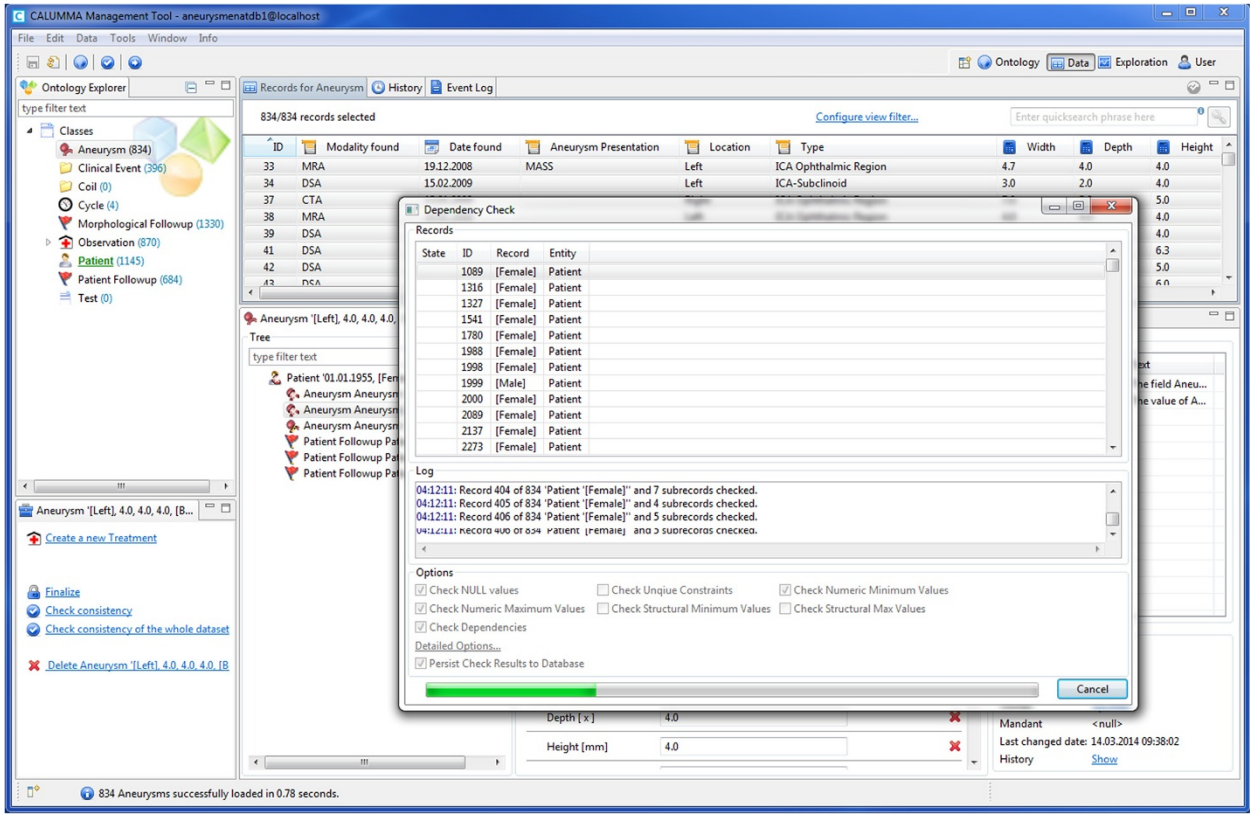

Figure 3 The data view of the Management Tool showing a currently running data validity check.

From a technical point of view, the additional level of abstraction also contains risks and complications. Data that can easily be queried out of an M_1 data-model, needs to be assembled by multiple joins out of the meta-model. This is not just an issue when talking about performance, it also complicates the source code. Software developers must get used to think in meta-model terms, which means dealing with another level of abstraction. Especially when it comes to queries meta-models can be very challenging. Another drawback that has to be accepted is the fact the data is distributed unevenly through the meta data-model. Although, real applications did not yet yield any performance issues, benchmarks and load tests indicated that abstraction overhead reduces the performance for very large data-sets (keyword big-data) in comparison to conventional, relational M_1 data models.

Conclusion

We critically revised the knowledge discovery process in the field of scientific medical research. By assuming, that the medical domain expert takes the central role in this process, a number of technical challenges appear to this expert. Consequently, an intelligent software system is needed, to actively support the domain expert in dealing with this challenges. Since every research project demands its very own data structures (or data-model or domain ontology), we followed a generic, ontology-centered approach. By definition of the actual domain ontology, the whole data warehouse adapts its appearance and behavior to this ontology — at run-time. By this means, the data warehouse appears to the user like a custom made solution for his domain data structures. It covers the process of knowledge discovery from the task of data modeling, over data acquisition, data integration and aggregation to exploratory data analysis in one integrative system. It has already been successfully applied in a number of medical research studies and clinical benchmarking projects.

Based on the first experiences with this approach we identified a number of research tasks for further development. The limited ability to process unstructured, free-text data is very unsatisfying considering the fact that clinical information systems mostly contains this type of data in form of doctor's letters, care instructions, and clinical findings. Although, efforts are known to increase the level of structuring in clinical information systems, it can be assumed that the situation we stay the same for the next years. So, one future research tasks will be the field of ontology-guided information extraction from un- or semi-structured data. Based on the idea of ontology-guided visual analytics we plan to integrate a number of non-linear data analysis, machine learning, and data-mining algorithms, to allow the user to identify non-linear correlations or clusters in his data.

Competing interests

The proposed system is sold commercially under the brand CALUMMA by RISC Software GmbH, Research Unit Medical-Informatics, Hagenberg, Austria, which employs three of the authors of this article (D. Girardi, J. Dirnberger, M. Giretzlehner). RISC Software GmbH is a non-profit organization owned and funded by the Johannes Kepler University Linz and the local Upper Austrian government and any earnings are reinvested into further development of CALUMMA. No money was paid by RISC Software GmbH for the use of CALUMMA.

Authors' contributions

All authors initially designed the basic principle of the generic, ontology-based research infrastructure. JD and DG were responsible for the implementation and DG is project-leader. The presented Management Tool was implemented by DG and JD was mainly responsible for the ontology-based web interface. The manuscript was drafted and written mainly by DG with contributions from JD and MG. All authors read and approved the final manuscript.

Acknowledgements

We want to thank the Department for Radiology of the Landesnervenklinik Wagner Jauregg under the administration of Dr. Johannes Trenkler for providing precious insights in everyday clinical research and a longstanding research cooperation. Furthermore, we want to thank the Department for Process Manamgent in Healthcare of the University of Applied Sciences in Steyr for a longstanding research cooperation and accompanying the system into the complex field of clinical benchmarking.
The presented work was partially funded by The Austrian Research Promotion Agency (FFG) and the Country of Upper Austria.

Received: 30 December 2014 Accepted: 21 February 2015
Published: 20 July 2015

References

1. Cios KJ, Pedrycz W, Swiniarski RW, Kurgan LA. Data Mining: A Knowledge Discovery Approach, 1st edn. New York: Springer; 2007. http://www.worldcat.org/isbn/0387333339.
2. Cios KJ, William Moore G. Uniqueness of medical data mining. Artificial intelligence in medicine. 2002;26(1):1–24.
3. Anderson NR, Lee ES, Brockenbrough JS, Minie ME, Fuller S, Brinkley J, Tarczy-Hornoch P. Issues in biomedical research data management and analysis: Needs and barriers. Journal of the American Medical Informatics Association. 2007;14(4):478–488. doi:10.1197/jamia.M2114. http://jamia.bmj.com/content/14/4/478.full.pdf+html.
4. Prokosch H-U, Ganslandt T. Perspectives for medical informatics. Methods Inf Med. 2009;48(1):38–44.
5. Inmon W. Building the Data Warehouse. New York: Wiley; 1993.
6. Frankel D. Model Driven Architecture: Applying MDA to Enterprise Computing. New York: Wiley; 2003.
7. Cruz AMR, Faria JP. A metamodel-based approach for automatic user interface generation In: Petriu D, Rouquette N, Haugen A, editors. Model Driven Engineering Languages and Systems. Lecture Notes in Computer Science, vol. 6394. Berlin Heidelberg: Springer; 2010. p. 256–270.
8. Renggli L, Ducasse S, Kuhn A. Magritte - a meta-driven approach to empower developers and end users In: Engels G, Opdyke B, Schmidt D, Weil F, editors. Model Driven Engineering Languages and Systems. Lecture Notes in Computer Science, vol. 4735. Berlin Heidelberg: Springer; 2007. p. 106–120.
9. Zavaliy T, Nikolski I. Ontology-based information system for collecting electronic medical records data. In: Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2010 International Conference On. Lviv: Publishing House of Lviv Polytechnic; 2010. p. 125.
10. Tran QD, Kameyama W. A proposal of ontology-based health care information extraction system: Vnhies. In: Research, Innovation and Vision for the Future, 2007 IEEE International Conference On. New Jersey: IEEE; 2007. p. 1–7.
11. Kataria P, Juric R, Paurobally S, Madani K. Implementation of ontology for intelligent hospital wards. In: Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, title = Implementation of Ontology for Intelligent Hospital Wards. New Jersey: IEEE; 2008. p. 253.

12. Kiong YC, Palaniappan S, Yahaya NA. Health ontology system. In: Information Technology in Asia (CITA 11), 2011 7th International Conference On. New Jersey: IEEE; 2011. p. 1–4.
13. Lozano-Rubí R, Pastor X, Lozano E. Owling clinical data repositories with the ontology web language. JMIR Medical Informatics. 2014;2(2):14.
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (redcap) - a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009;42(2):377–381.
15. Kurgan LA, Musilek P. A survey of knowledge discovery and data mining process models. The Knowledge Engineering Review. 2006;21(01):1–24.
16. Meta Object Facility (MOF) Specification. Object Management Group. OMG-Document ad/97-08-14. http://www.omg.org/cgi-bin/doc?ad/97-08-14.pdf.
17. Chen PP-s. The entity relationship model - toward a unified view of data. ACM Transactions on Database Systems. 1976;1(1):9–36.
18. Girardi D, Arthofer K. An ontology-based data acquisition infrastructure - using ontologies to create domain-independent software systems. In: KEOD 2012 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4 - 7 October, 2012. Barcelona: SciTePress; 2012. p. 155–160. doi:10.5220/0004108101550160.
19. Girardi D, Dirnberger J, Trenkler J. A meta model-based web framework for domain independent data acquisition. In: ICCGI 2013, The Eighth International Multi-Conference on Computing in the Global Information Technology. Nice, France: International Academy, Research, and Industry Association; 2013. p. 133–138.
20. Roddick JF, Fule P, Graco WJ. Exploratory medical knowledge discovery: experiences and issues. SIGKDD Explor. Newsl. 2003;5(1):94–99. doi:10.1145/959242.959243.
21. Free On-Line Dictionary of Computing. Free On-Line Dictionary of Computing - Expression. http://foldoc.org/Expression. http://foldoc.org/Expression.
22. Girardi D, Küng J, Giretzlehner M. A meta-model guided expression engine. In: Intelligent Information and Database Systems. Cham Heidelberg New York Dordrecht London: Springer; 2014. p. 1–10.
23. Behrens JT, Yu C-H. New Jersey: John Wiley & Sons, Inc.; 2003. doi:10.1002/0471264385.wei0202.
24. Thomas JJ, Cook KA. A visual analytics agenda. Computer Graphics and Applications, IEEE. 2006;26(1):10–13. doi:10.1109/MCG.2006.5.
25. Kohonen T. Self-Organizing Maps - Third Edition. Berlin Heidelberg New York: Springer; 2001.
26. Inselberg A, Dimsdale B. Parallel coordinates. In: Human-Machine Interactive Systems. US: Springer; 1991. p. 199–233.
27. Sammon JW. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers. 1969;18:401–409.
28. Girardi D, Giretzlehner M, Küng J. Using generic meta-data-models for clustering medical data. In: ITBAM. Heidelberg Dordrecht London New York: Springer; 2012. p. 40–53.
29. Girardi D, Arthofer K. Supporting knowledge discovery in medicine. Studies in health technology and informatics. 2013;198:147–155.

doi:10.1186/2056-5917-1-6
Cite this article as: Girardi *et al.*: An ontology-based clinical data warehouse for scientific research. *Safety in Health* 2015 **1**:6.